Sankhyā : The Indian Journal of Statistics
2000, Volume 62, Series A, Pt. 1, pp. 40-48

AN ALTERNATIVE TEST ON BERNOULLI SUCCESS PROBABILITY

By UTTAM BANDYOPADHYAY
University of Calcutta, Calcutta
and
ATANU BISWAS
Indian Statistical Institute, Calcutta

SUMMARY. This paper provides a test procedure for the problem of testing on Bernoulli success probability in case of costly trials. The proposed test is based on a sampling scheme which we call 'randomized play-the-looser' rule. Some exact and asymptotic results related to the test are studied.

1. Introduction

The paper considers the old problem of testing on Bernoulli success probability which can be set in the following way: Suppose there is a sequence of independent Bernoulli trials with outcomes $Y_1, Y_2, \cdots$, where Y_i takes the values 1 (success for the i-th trial) with unknown probability p and 0 (failure for the i-th trial) with probability $q = 1 - p$. Then for a prefixed value p_0, the problem is to test

$$H : p = p_0 \quad \text{against} \quad H_a : p > p_0. \qquad \ldots (1.1)$$

One of the existing approaches for solving (1.1) is to consider n sample observations $Y_1, Y_2, \cdots, Y_n$ and find $U = \sum_{i=1}^{n}(1 - Y_i)$, the total number of failures in n trials. Then a left-tailed test based on U is appropriate and is given by:

$$\text{Reject } H \text{ if } U < c \text{ and accept } H \text{ otherwise,} \qquad \ldots (1.2)$$

where 'c' is a positive integer so chosen that the level of the test is γ (prefixed). Suitable randomization can be done to make it a size γ test. The test is uniformly most powerful (UMP).

In many practical situations the trials involve some other costs except the sampling cost like monetary cost for equipments of a machine, ethical cost in clinical trials. Consequently such a costly trial necessitates the use of a test procedure with smallest possible sample size when an alternative is true.

Paper received. October 1996; revised June 1999.

AMS (1991) *subject classification.* Primary 62F03; secondary 62L05.

Key words and phrases. Randomized play-the-looser rule; Urn model; Power per unit sample; Asymptotic power.

For comparing two treatments in a clinical trial, Zelen (1969), assuming sequential entrance, introduced the concept of 'play-the-winner' rule for dichotomous responses of patients. As a modification towards Zelen's approach, Wei and Durham (1978) and Wei (1979) introduced the idea of 'randomized play-the-winner' (RPW) rule. For both the rules the goal was to allocate a larger number of patients to the better treatment.

For testing (1.1) we modify the RPW rule as follows: First introduce a hypothetical treatment with success probability p_0. We call it treatment A. Suppose the Bernoulli trials under consideration are responses by treatment B. As, under an alternative the success probability (p) of treatment B is greater than that of treatment A, we call treatment A 'a hypothetical looser'. We need not have the existence of such a looser. If a trial is needed from treatment A, we simply toss a coin with probability p_0 of occurrence of 'head' (success). Then, with the help of such hypothetical treatment, we introduce a sampling scheme called randomized play-the-looser (RPL) rule. Let us illustrate the rule by an urn model as follows:

Start with an urn having two types of balls A and B, $(k-1)\alpha$ balls of type A ($k \geq 2$) and α balls of type B. For any trial we consider it from either treatment by drawing a ball from the urn with replacement. If a ball of kind A is obtained we toss a coin with success probability p_0. On the other hand, if a ball of kind B is obtained, we take a trial from the treatment under consideration assuming instantaneous response. If a success by treatment B is obtained we add an additional $(k-1)\beta$ balls of kind A, and if a failure by treatment B is obtained we add an additional $(k-1)\beta$ balls of kind B in the urn. On the other hand, a failure by treatment A results the addition of $(k-1)\beta$ balls of kind A and a success by treatment A results the addition of $(k-2)\beta$ balls of kind A along with β balls of kind B. For a given (α, β), this is abbreviated as RPL(α, β) rule. Thus, when the alternative is true, an RPL rule results more trials from the hypothetical treatment A involving only the trouble of tossing a coin of zero cost.

In section 2, the proposed RPL rule is used to provide an alternative test for the problem (1.1) with a rationale. There we see that the expected number of trials corresponding to treatment B decreases as p increases. It is also shown that the proposed test is an improvement over the existing test with respect to the criterion called 'power per unit sample' (PPUS). In section 3, some asymptotics related to the proposed test are established. These are, respectively, the null and non-null distributions of the test statistic, the consistency of the test, and the limiting proportion of experimental units receiving the treatment B. Finally, section 4 ends with a discussion.

2. Proposed Test Procedure and Related Exact Results

Suppose there is a sequential chain of patient's entrance upto a maximum of n patients. First, corresponding to the i-th entering patient we set a pair $\{\delta_i, Z_i\}$ of indicator variables defined by:

$\delta_i = 1$ or 0 according as treatment A or B applied using RPL(α, β) procedure;

$Z_i = 1$ or 0 according as success or failure occurs for the i-th patient.

Then, we can observe a pair of random variables defined by

$N_n = \sum_{i=1}^{n}(1-\delta_i)$ = Number of patients treated by treatment B, and
$T_n = \sum_{i=1}^{n}(1-Z_i)(1-\delta_i)$ = Number of failures by treatment B.

Hence $S_n = T_n/N_n$ gives the proportion of failures by treatment B, which is expected to be larger under H than under H_a.

We start with the problem of finding a most powerful test by using the Neyman-Pearson lemma. The probability functions corresponding to H_a and $p = p_0$ for the i-th sample observation are respectively given by

$$\{p_0 Z_i + (1-p_0)(1-Z_i)\}^{\delta_i}\{pZ_i + (1-p)(1-Z_i)\}^{1-\delta_i},$$

and

$$\{p_0 Z_i + (1-p_0)(1-Z_i)\}.$$

Hence, conditioning at every stage, the joint probability functions corresponding to H_a and H are respectively

$$f_{p_0}(Z_1,\cdots,Z_n;\delta_1,\cdots,\delta_n) = \prod_{i=1}^{n}\{(1-p_0)+(2p_0-1)Z_i\}, \qquad \ldots(2.1)$$

$$f_p(Z_1,\cdots,Z_n;\delta_1,\cdots,\delta_n) = \prod_{i=1}^{n}\{(1-p_0)+(2p_0-1)Z_i\}^{\delta_i}\{(1-p)+(2p-1)Z_i\}^{1-\delta_i}. \qquad \ldots(2.2)$$

Then, for any k_0 (>0), a most powerful test is given by

$$\begin{aligned}\omega \;=\; \{(Z_1,\cdots,Z_n;\delta_1,\cdots,\delta_n) : f_p(Z_1,\cdots,Z_n;\delta_1,\cdots,\delta_n) > \\ k_0 f_{p_0}(Z_1,\cdots,Z_n;\delta_1,\cdots,\delta_n)\}.\end{aligned} \qquad \ldots(2.3)$$

Then after some routine steps the statement under (2.3) reduces to

$$b_1\sum_{i=1}^{n}(1-\delta_i) - b_2\sum_{i=1}^{n}(1-Z_i)(1-\delta_i) > k_1, \qquad \ldots(2.4)$$

where $b_1 = \log_e(p/p_0)$, $b_2 = \log_e\left(\frac{p(1-p_0)}{p_0(1-p)}\right)$ and $k_1 = \log_e k_0$. Thus we get a class of tests determined by the critical region:

$$a_1\sum_{i=1}^{n}(1-\delta_i) - a_2\sum_{i=1}^{n}(1-Z_i)(1-\delta_i) > a_3, \qquad \ldots(2.5)$$

by varying a_1, a_2 and a_3. Now, from (2.4), we get, provided $\sum_{i=1}^{n}(1-\delta_i) > 0$,

$$\sum_{i=1}^{n}(1-Z_i)(1-\delta_i) \Big/ \sum_{i=1}^{n}(1-\delta_i) < b_1/b_2 - k_1 \Big/ \left(b_2\sum_{i=1}^{n}(1-\delta_i)\right). \qquad \ldots(2.6)$$

If $\sum_{i=1}^{n}(1-\delta_i)=0$, then H is automatically rejected. For operational convenience and easy interpretation of the test, we ignore the random part in the right hand side of (2.6). This motivates us to use S_n as our test statistic and a left-tailed test would be appropriate. Suppose, for a preassigned $\gamma \in (0,1)$, we have a non-negative integer c_1 satisfying

$$P_H\{S_n \le c_1\} \le \gamma < P_H\{S_n \le c_1+1\}. \qquad \ldots (2.7)$$

Then we reject H at the level γ iff

$$S_n \le c_1, \qquad \ldots (2.8)$$

and accept H otherwise. The test can be suitably randomized to get a size γ test.

Let p_{i+1} be the conditional probability that $\delta_{i+1}=1$ given all the previous assignments $\delta_1,\cdots,\delta_i$ and responses $Z_1,\cdots,Z_i$. Then it can be easily seen that

$$\begin{aligned} p_{i+1} &= \left\{(k-1)\alpha+\beta\left((k-1)\sum_{j=1}^{i}\delta_j+(k-1)\sum_{j=1}^{i}Z_j\right.\right. \\ &\quad \left.\left.-k\sum_{j=1}^{i}\delta_j Z_j\right)\right\} \Big/ (k\alpha+i(k-1)\beta), \end{aligned} \qquad \ldots (2.9)$$

and hence, the marginal distributions of δ_i's can be obtained by the method of induction as

$$P(\delta_{i+1}=1)=\frac{k-1}{k}+d_{i+1}, \qquad \ldots (2.10)$$

where $d_1=0$, and for $i \ge 1$,

$$\begin{aligned} d_{i+1} &= \frac{i\beta}{k\alpha+i(k-1)\beta}\frac{k-1}{k}(p-p_0)+\frac{(k-1)\beta}{k\alpha+i(k-1)\beta}(1+p_0-p)\sum_{j=1}^{i}d_j \\ &\quad -\frac{kp_0\beta}{k\alpha+i(k-1)\beta}\sum_{j=1}^{i}d_j. \end{aligned} \qquad \ldots (2.11)$$

Note that (2.9)-(2.11) show that the procedure depends on α and β only through β/α. Under alternative, the distribution of Z_{i+1} depends on δ_{i+1} and hence on all the previous (δ_j, Z_j)'s. But, under H, the Z_i's are independently and identically distributed (i.i.d.) Bernoulli (p_0) random variables, and δ_i's are identically distributed Bernoulli $(1-k^{-1})$ random variables but not independent.

From (2.10)-(2.11), it is clear that the probability of obtaining a ball of kind A is $(1-k^{-1})$ under H. But, under an alternative, the probability is $(1-k^{-1})$ for the first draw, and is greater than $(1-k^{-1})$ from the second draw onwards. The proposed test, compared to the existing test with n_0 observations, requires

$n = kn_0$ experimental units of which some correspond to B and the remaining to A. From the probability distributions given above, the expected number of allocations to treatment B (which is the treatment under consideration) is $E(\sum_{i=1}^{n} \delta_i) = \sum_{i=1}^{n}(\frac{1}{k} - d_i) = n_0 - \sum_{i=1}^{n} d_i$, which is less than n_0 under an alternative and equal to n_0 under H. Hence, as k increases, we get a better test which is also clear from Table 2.1. But this involves more time for tossing a coin and drawing balls from the urn with little extra cost.

Here, for the proposed test procedure, we are mainly interested in two performance characteristics, viz., the power function ($P(p)$) and, for treatment B, the average sample number ($A(p)$) required to perform the test. It is to be noted that the expected number of coin tossings required is given by $n - A(p)$. The main goal of the proposed sampling scheme is to have a fewer number of patients when an alternative is true, i.e., to arrive at a decision by exposing a fewer number of such costly trials. But, as the reduction of sample size results in reduction of power, we consider $P(p)/A(p)$, which is the PPUS, as our basic criterion and compare it with that of the corresponding test (1.2) based on a sample size n/k. It can be easily shown that $A(p_0) = n/k$. In the following table, taking $p_0 = 0.5$ and $n = 15k$, $k = 2, 3, 4, 5$, we compute the performance characteristics of the proposed and the existing tests at some selected values of p. Here $A(p)$ values are exact, but $P(p)$ values are obtained by 10000 simulations. Figure 1 also shows the comparison between the PPUS of the proposed test and that of the existing one.

Table 2.1. Performance characteristics of the proposed test and the existing test.

p	Existing test	$k = 2$	$k = 3$	$k = 4$	$k = 5$
		0.0500	0.0500	0.0500	0.0500
0.5		15.0000	15.0000	15.0000	15.0000
	0.0033	0.0033	0.0033	0.0033	0.0033
		0.1592	0.1525	0.1560	0.1529
0.6		13.8759	13.5290	13.3448	13.2269
	0.0126	0.0115	0.0113	0.0117	0.0116
		0.3849	0.3667	0.3644	0.3455
0.7		12.8844	12.2788	11.9677	11.7729
	0.0311	0.0299	0.0299	0.0304	0.0293
		0.6965	0.6645	0.6775	0.6415
0.8		12.0065	11.2105	10.8139	10.5705
	0.0529	0.0580	0.0593	0.0627	0.0609
		0.9418	0.9203	0.9147	0.9109
0.9		11.2269	10.2926	9.8402	9.5677
	0.0652	0.0839	0.0894	0.0930	0.0952
		0.9930	0.9836	0.9825	0.9755
0.95		10.8697	9.8818	9.4099	9.1281
	0.0666	0.0914	0.0995	0.1044	0.1069
		0.9998	0.9992	0.9990	0.9990
0.99		10.5981	9.5738	9.0895	8.8082
	0.0667	0.0943	0.1044	0.1099	0.1135

The three entries in each cell correspond to $P(p)$, $A(p)$ and PPUS at p respectively.

The above table shows that PPUS of the proposed test is almost same as that of the existing one for all values of $p \leq 0.75$. But there is a gain in PPUS for the proposed test at all $p > 0.75$.

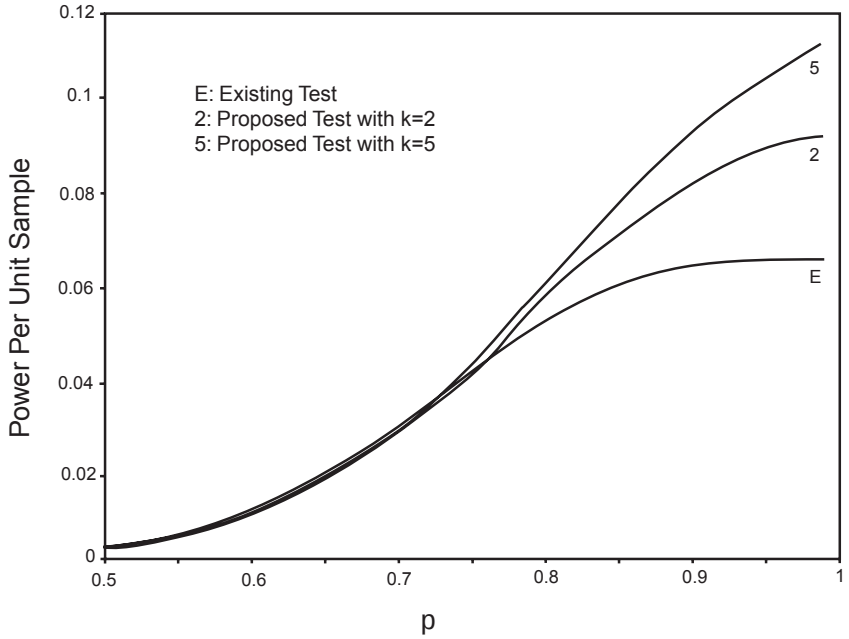

Figure 1. PPUS of the existing and proposed tests.

3. Some Asymptotic Results

As in Bandyopadhyay and Biswas (1996) the following lemma can be proved.

LEMMA 3.1. *There exists* $\mu^* \in (0,1)$ *such that, as* $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i) \xrightarrow{P} \mu^*. \qquad \ldots (3.1)$$

Now we have the following theorem:

THEOREM 3.1. *Under* H, *as* $n \to \infty$,

$$n^{1/2}(S_n - q_0) \xrightarrow{d} N(0, \sigma^2),$$

where $\sigma^2 = kq_0(1-q_0)$, and $q_0 = 1 - p_0$.

PROOF. Note that, as in Lemma 4.1, we have

$$N_n/n \xrightarrow{P} \frac{1}{k}, \quad \text{as} \quad n \to \infty, \qquad \ldots (3.2)$$

and hence

$$n^{1/2}(S_n - q_0) = \frac{n^{1/2}}{N_n}(T_n - N_n q_0) \overset{P}{\simeq} \frac{k}{\sqrt{n}}(T_n - N_n q_0), \qquad \ldots (3.3)$$

where '$X_n \overset{P}{\simeq} Y_n$' means that, as $n \to \infty$, X_n/Y_n tends to 1 in probability. Denoting by $(i_1, \cdots, i_{N_n})$ the set of random indices for which $\delta_i = 0$, we have, under H,

$$\begin{aligned} & \frac{1}{\sqrt{n}}(T_n - N_n q_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1-\delta_i)(1 - Z_i - q_0) = \frac{1}{\sqrt{n}} \sum_{i \in (i_1, \cdots, i_{N_n})} (1 - Z_i - q_0) \\ \overset{d}{\to} & \frac{1}{\sqrt{n}} \sum_{i=1}^{N_n} (1 - Z_i - q_0) \\ = & \frac{1}{\sqrt{n}} \sum_{i=1}^{\nu} (1 - Z_i - q_0) + \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{N_n} (1 - Z_i - q_0) - \sum_{i=1}^{\nu} (1 - Z_i - q_0) \right], \end{aligned} \quad \ldots (3.4)$$

where $\{\nu = \nu(n)\}$ be a sequence such that $\nu \to \infty$ and $\nu/n \to 1/k$ as $n \to \infty$. As Z_i's are i.i.d. Bernoulli (p_0), the first part of the right hand side of (3.4) converges in distribution to $N(0, q_0(1-q_0)/k)$ and, by Kolmogorov's inequality (see Billingsley (1979, pp. 320-321)), the second part of (3.4) converges in probability to zero. Combining all these the theorem follows. □

Now, using the above theorem, the proposed test can be approximated by: Reject H at the level $\gamma \in (0,1)$ iff

$$S_n < q_0 - n^{-1/2}(k q_0(1-q_0))^{1/2} \Phi^{-1}(1-\gamma), \quad \ldots (3.5)$$

where $\Phi^{-1}(1-\gamma)$ is the $(1-\gamma)$ percentile point of an $N(0,1)$ distribution. The test is then asymptotically size γ.

To find the limiting value of the power function we prove the following theorem:

THEOREM 3.2. *Under* H_a, $P(p) \to 1$ *as* $n \to \infty$.

PROOF. By Lemma 3.1, we have

$$N_n \overset{P}{\simeq} \nu_n^*, \quad \ldots (3.6)$$

where it is always possible to have a sequence of positive integers $\{\nu_n^*\}$ such that $\nu_n^* \to \infty$, and $\nu_n^*/n \to \mu^*$ as $n \to \infty$. Hence

$$S_n = T_n/N_n \overset{P}{\simeq} T_n/\nu_n^* = S_n^* \quad \text{(say)}. \quad \ldots (3.7)$$

It is easy to see that, given N_n, the conditional distribution of T_n is binomial (N_n, q), where $q = 1-p$, and it is easy to show that as $n \to \infty$, $E(S_n^*)$ and $V(S_n^*)$ converge respectively to q and 0, which by (3.5), implies

$$S_n \overset{P}{\to} q.$$

Then, using (3.5) the theorem follows. □

As the proposed test is consistent the asymptotic power can be obtained only from the limiting distribution of the test statistic under a sequence of local alternatives : $p = p_n = p_0 + c/\sqrt{n}$, $c > 0$, $n \geq 1$. It can be easily argued that under $\{p_n\}$, as $n \to \infty$,

$$\sqrt{n}\left(\frac{T_n}{N_n} - q_0\right) \xrightarrow{d} N(-c, kq_0(1-q_0)), \qquad \ldots (3.8)$$

and hence the asymptotic power is given by

$$\Phi\left[\Phi^{-1}(1-\gamma) + \frac{c}{\sqrt{kq_0(1-q_0)}}\right]. \qquad \ldots (3.9)$$

The initial goal of our sampling design was to have a fewer number of patients under alternative. (Table 2.1 supports our assertion.) As for any $p > p_0$, the sequence $\{d_i,\ i \geq 2\}$ is monotonic and bounded, writing $\lim_{i\to\infty} d_i = d$, we have, by Toeplitz's lemma, as $n \to \infty$,

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\delta_i\right) = \frac{k-1}{k} + \frac{1}{n}\sum_{i=1}^{n} d_i \to \frac{k-1}{k} + d, \qquad \ldots (3.10)$$

as the limiting proportion of coin tosses, where by (2.11), we have

$$d = \frac{1}{k}(p - p_0) + (1 + p_0 - p)d - \frac{k}{k-1}p_0 d, \qquad \ldots (3.11)$$

yielding $d = (k-1)(p-p_0)/[k((k-1)p + p_0)]$. Hence the limiting value of the expected proportion of trials that are allocated to treatment B is

$$1 - \lim_{i\to\infty} E\left(\frac{1}{n}\sum_{i=1}^{n}\delta_i\right) = \frac{p_0}{(k-1)p + p_0}. \qquad \ldots (3.12)$$

Note that this limiting proportion is free of any choice of (α, β).

4. Concluding Remarks

For large sample size, the test procedure is independent of any choice of β/α, but the small sample behaviour depends on its choice. So it is the experimenter's task to choose β/α compromising between power and ASN.

It is to be noted that an early stopping can easily be employed to our present sampling scheme. This will retain the same power at the expense of possibly smaller sample size. But this is a routine extension of any sequential decision making procedure, and hence we have not studied it. The proposed test, being a terminal one, is thus comparable with an existing fixed sample size procedure. If an early stopping were done, we could compare the test with an SPRT-type procedure with truncation. But such a study is not undertaken simply because the aim of the present paper is to establish the usefulness of the proposed RPL sampling scheme for costly trials.

Acknowledgment. The authors wish to thank the referee and the editorial board for their valuable comments which led to an improvement over some earlier versions of the manuscript.

References

BANDYOPADHYAY, U. AND BISWAS, A. (1996). Delayed response in randomized play-the-winner rule: a decision theoretic outlook. *Calcutta Statistical Association Bulletin* **46**, 69-88.

BILLINGSLEY, P. (1979). *Probability and Measure*, John Wiley and Sons, New York.

WEI, L. J. (1979). The generalized Polya's urn for sequential medical trials. *Ann. Statist.* **7** , 291-296.

WEI, L. J. AND DURHAM, S. (1978). The randomized play-the-winner rule in medical trials. *J. Am. Statist. Assoc.* **73** , 838-843.

ZELEN, M. (1969). Play-the-winner rule and the controlled clinical trial. *J. Am. Statist. Assoc.* **64** , 131-146.

UTTAM BANDYOPADHYAY
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALCUTTA
35 BALLYGUNGE CIRCULAR ROAD
CALCUTTA - 700 019
INDIA

ATANU BISWAS
APPLIED STATISTICS UNIT
INDIAN STATISTICAL INSTITUTE
203 B. T. ROAD
CALCUTTA - 700 035, INDIA
e-mail : atanu@isical.ac.in